RSAC 2025 Conference

Many Voices.
One Community.

SESSION ID: ANI-R02

The Telltale Signs of AI-Generated Emails: Building a Detection Engine

Rob Juncker

Chief Product Officer
Mimecast
/robjuncker

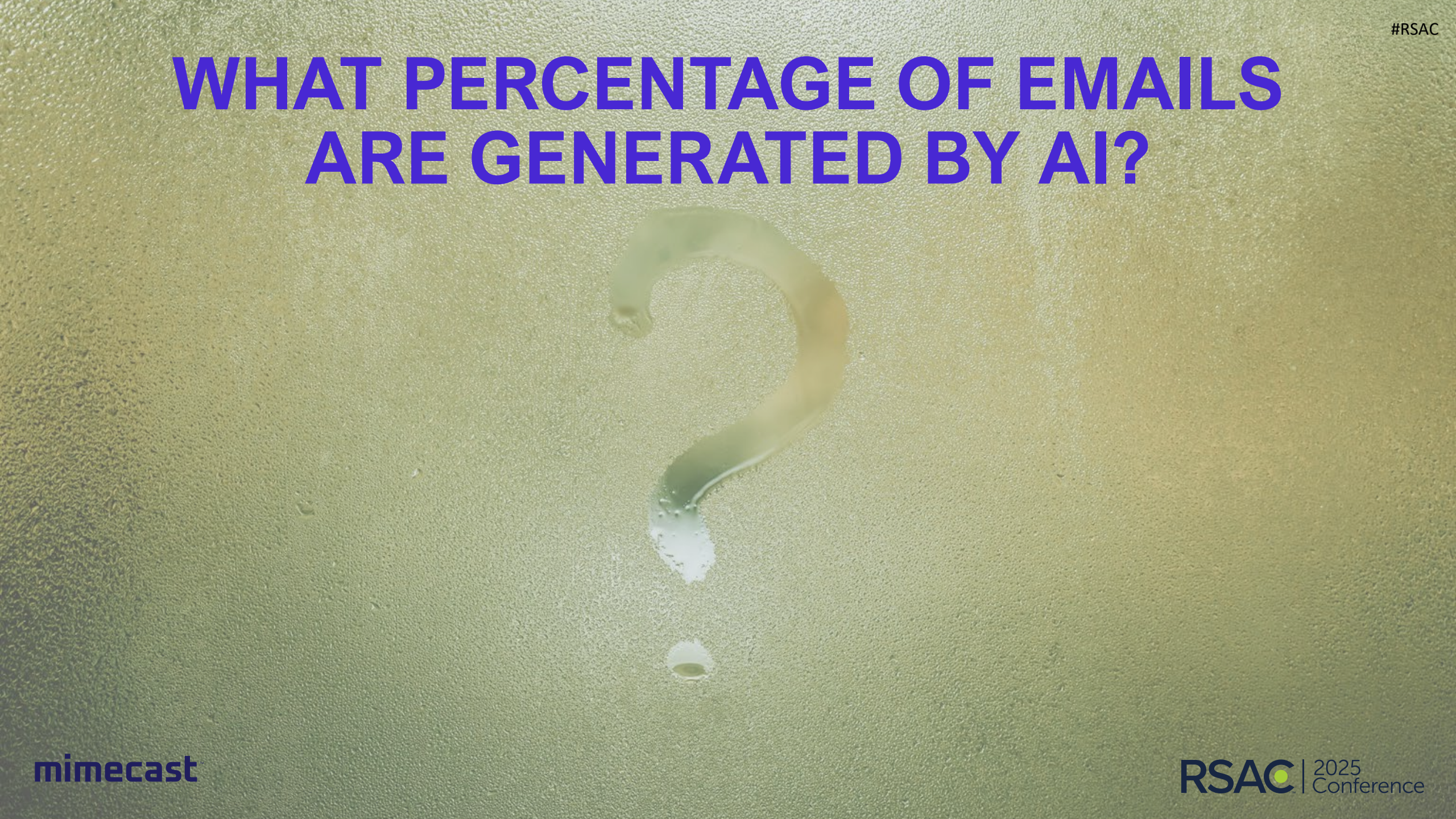
Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

WHAT PERCENTAGE OF EMAILS ARE GENERATED BY AI?



WHAT THE NEWS ARTICLES SAY

How AI is making phishing attacks more dangerous

Cybercriminals are using A
sophisticated business en
practitioners must fight fir

Sharon Shea, Executive Edito
Ashwin Krishnan, StandOutir

Published: 22 Oct 2024

HOME » SECURITY

SECURITY

Understanding Customized Phishing Emails in the Age of Generative AI

Hyperpersonalized email scams are surging. Here's how businesses can navigate them.

by J.P. Pressley

J.P. Pressley is a contributing writer to BizTech magazine and an editor at Manifest.

HBR

AI And Machine Learning

AI Will Increase the Quantity — and Quality — of Phishing Scams

by Fredrik Heiding, Bruce Schneier and Arun V

May 30, 2024

FAST COMPANY SUBSCRIBE

04-08-2024 | FAST COMPANY EXECUTIVE BOARD

The growing threat of AI in social engineering: How business can mitigate risks

AI social engineering is just getting started. Employers must make employees aware of these risks and train them to exercise their security intuition. Here's how.

SUBSCRIBE

RISK & COMPLIANCE JOURNAL

GenAI Increasingly Powering Scams, Wall Street Watchdog Warns

More swindlers are using the technology to trick firms and customers, Finra says

By Richard Vanderford Follow

Jan. 28, 2025 6:11 pm ET

by John Leyden
Senior Writer

Top 5 ways attackers use generative AI to exploit your systems

Feature

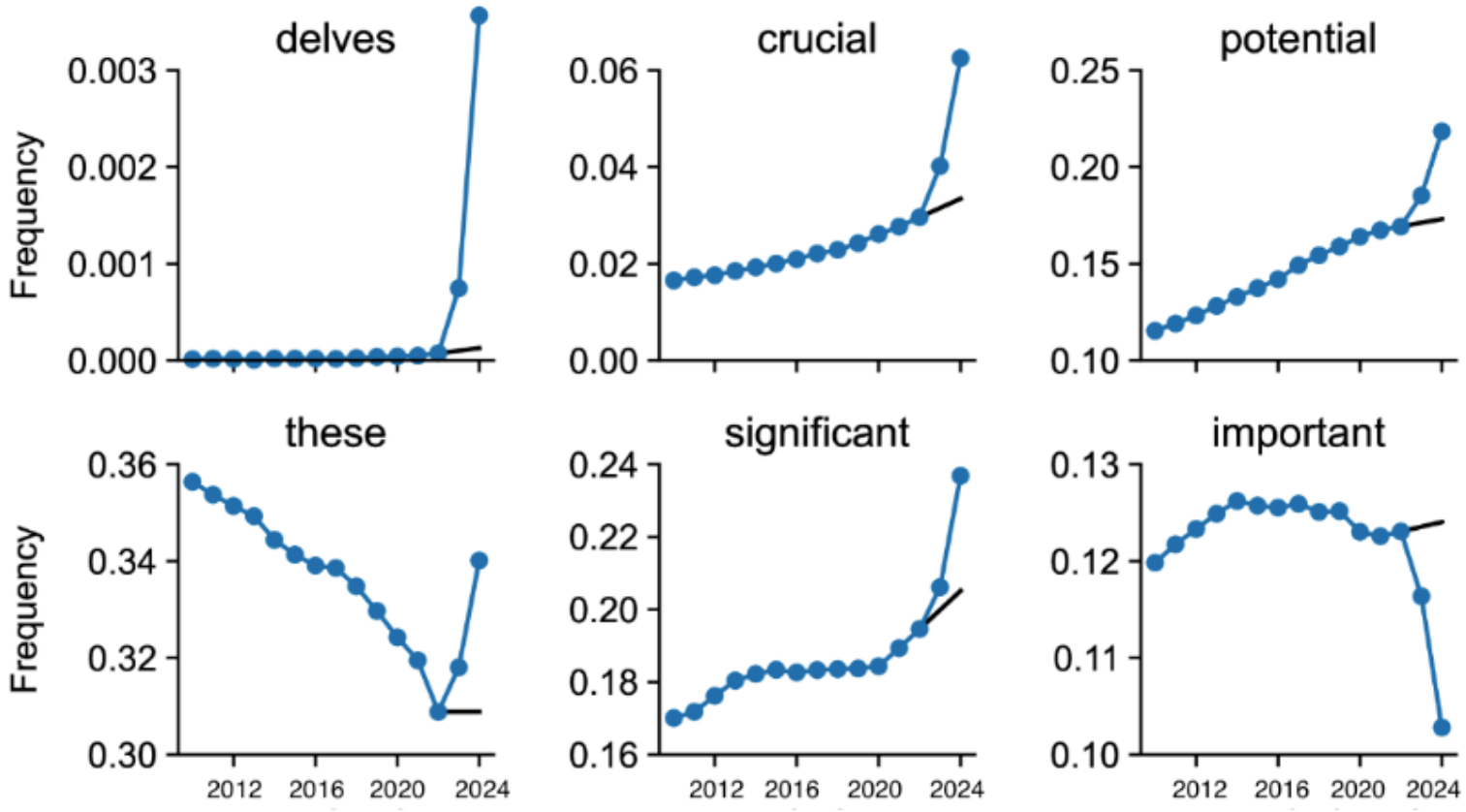
11 Feb 2025 • 8 mins

Browser Security Under Siege: The Alarming Rise of AI-Powered Phishing

Analysis reveals a 140% increase in browser phishing, including a 130% increase in zero-hour phishing attacks.

By Kevin Townsend
March 20, 2025

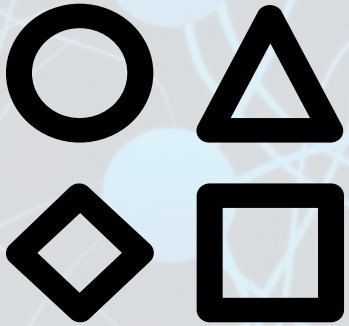
AI LANGUAGE HAS SOME TELLS



Delving into ChatGPT usage in academic writing through excess vocabulary – Cornell University

LLM DETECTION ENGINE

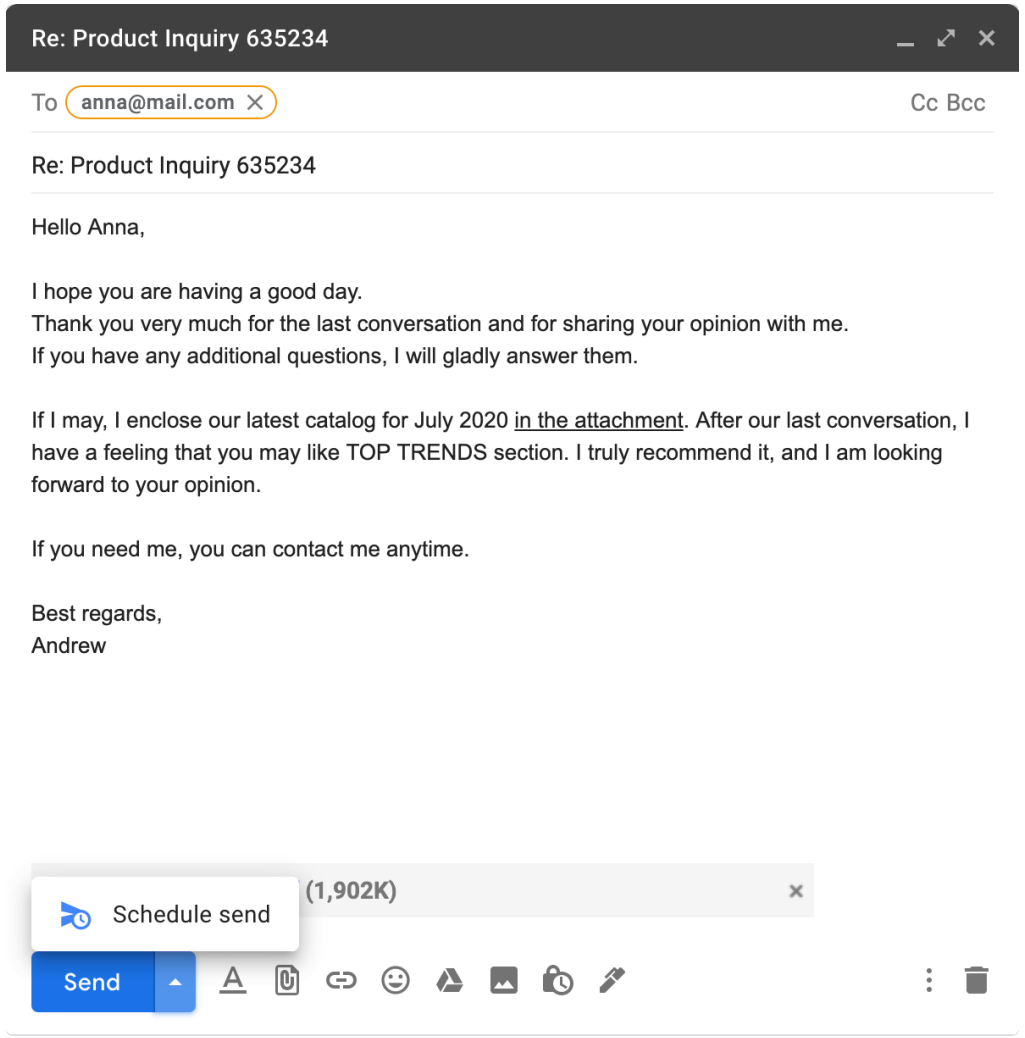
70K Training Data Set



Trained on Human and synthetic data

Testing on 17K collected emails

TRAINING THE MODEL

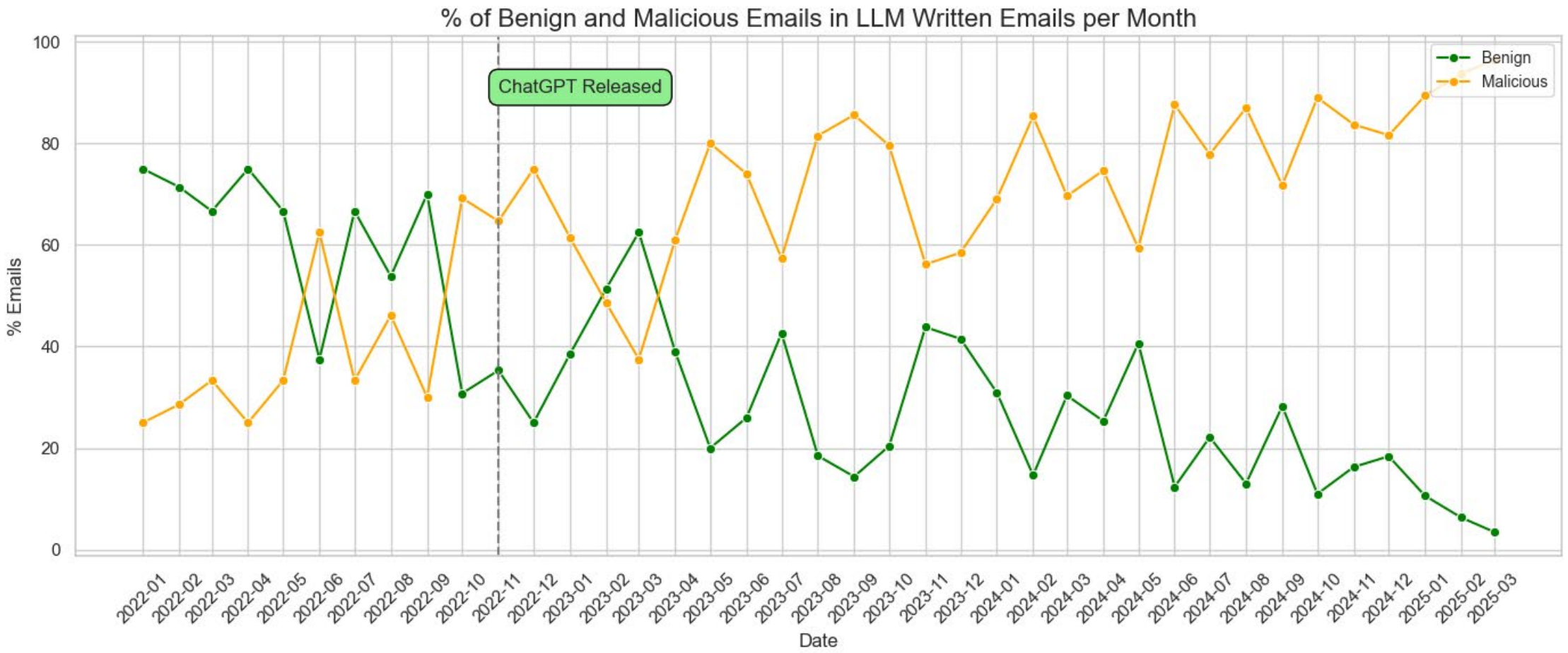

Re: Product Inquiry 635234
To anna@mail.com
Cc Bcc
Re: Product Inquiry 635234
Hello Anna,
I hope you are having a good day.
Thank you very much for the last conversation and for sharing your opinion with me.
If you have any additional questions, I will gladly answer them.
If I may, I enclose our latest catalog for July 2020 in the attachment. After our last conversation, I have a feeling that you may like TOP TRENDS section. I truly recommend it, and I am looking forward to your opinion.
If you need me, you can contact me anytime.
Best regards,
Andrew
(1,902K)
Schedule send
Send

- **Task:** Create a complete email based on summary, tone, and specified word length.
- **Content Creation:** Expand on the summary with original, contextually appropriate details.
- **Authenticity:** Use realistic, fictional details; avoid placeholders for names and contact info.
- **Formatting:** Write in plain text, no subject lines, and end with "===FINISHED===".
- **Restrictions:** No incomplete information, generic names, or process explanations allowed.

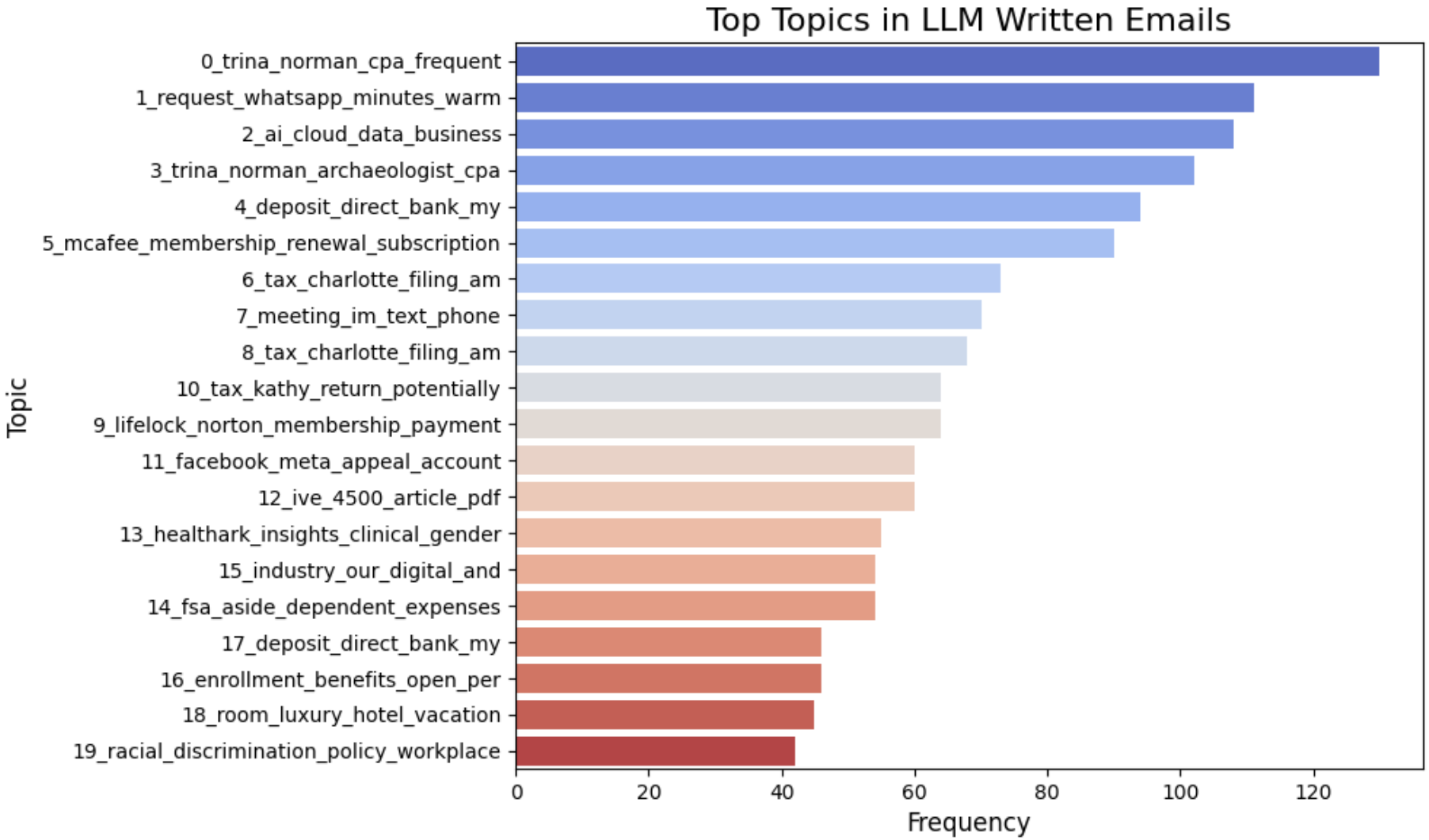
WHAT WE DISCOVERED

But this left us more questions…

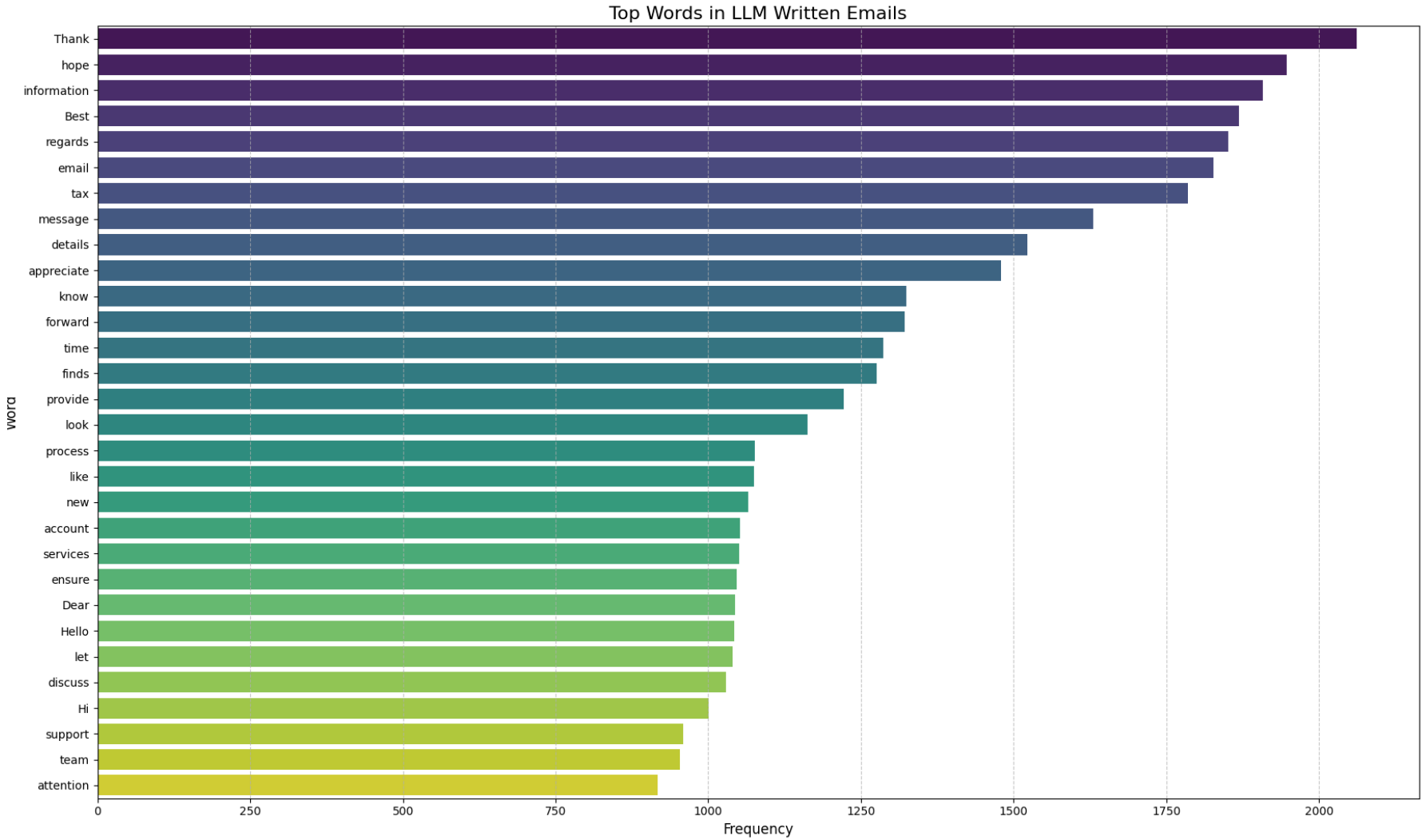
% OF AI WRITTEN EMAILS THAT ARE MALICIOUS



WHAT ARE MOST OF THE TOPICS FOCUSED ON?



WHAT WORDS ARE USED MOST OFTEN?



EXAMPLE #1 OF GEN AI PHISHING MESSAGE

Hi barry,

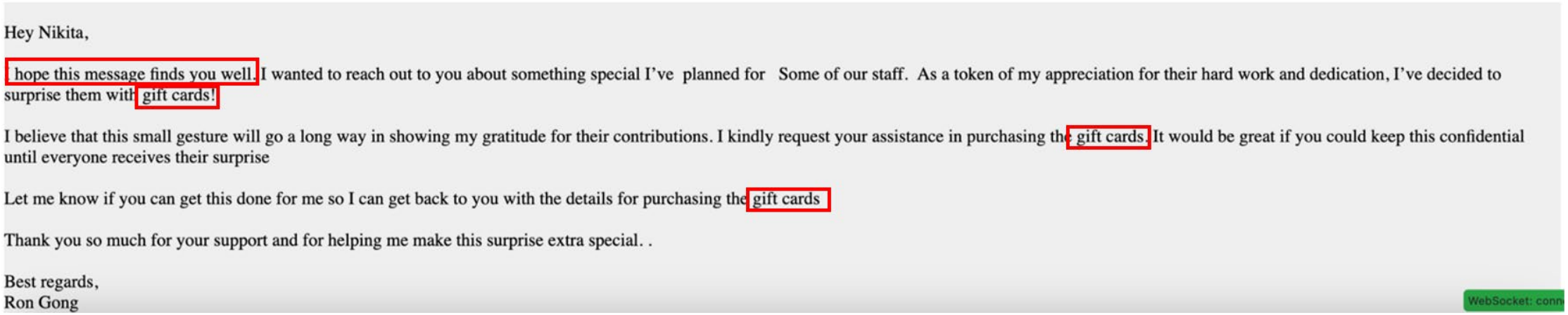
We hope this letter finds you well. Attached to this email is the Ginkgo Bioworks open benefits enrollment plan for your completion per Ginkgo Bioworks requests for all employees.

We wanted to remind you that Ginkgo Bioworks Open Benefits Enrollment plan period is coming to a close soon. If you haven't already done so, please take a moment to review your Ginkgo Bioworks benefits enrollment package and make any necessary changes per the attached document, before the deadline June 27

Thank you for your attention to this matter, and we look forward to your active participation in the open benefits enrollment process.

Ginkgo Bioworks

EXAMPLE #2 OF GEN AI BEC MESSAGE



Hey Nikita,

I hope this message finds you well. I wanted to reach out to you about something special I've planned for Some of our staff. As a token of my appreciation for their hard work and dedication, I've decided to surprise them with gift cards!

I believe that this small gesture will go a long way in showing my gratitude for their contributions. I kindly request your assistance in purchasing the gift cards. It would be great if you could keep this confidential until everyone receives their surprise

Let me know if you can get this done for me so I can get back to you with the details for purchasing the gift cards

Thank you so much for your support and for helping me make this surprise extra special. .

Best regards,
Ron Gong

WebSocket: conn

EXAMPLE #3 OF GEN AI PHISHING MESSAGE

Hey dhgdhgd

I stumbled upon your website while conducting some ad research, and I must say, I'm impressed with what I saw! Your design is sleek, navigation is intuitive, and your content is engaging—it's clear you've put a lot of effort into it.
However, as someone passionate about helping businesses thrive online, I noticed a few areas where some tweaks could make a big difference. For instance, Your website authority score is on average which is still not good enough and your mobile rsponsiveness needs to be attended to.

Rest assured, I'm here solely to help you succeed. No strings attached. My goal is to see your business boom, and I believe addressing these issues could take your website to the next level. I'd love to discuss these recommendations further and explore how we can work together to achieve your end goals. What do you think? Shall we schedule a quick meeting to delve deeper into this?

Attached below are documents and links showing some issues regarding your website, Kindly take a look to see what I am saying about your website.
hxxps://shorturl[.]at/mzALY

Looking forward to hearing from you and collaborating on enhancing your website's impact!
Best regards,
Sparklepulse Pro

RSAC 2025 Conference

How do you combat this problem?

You need integrated protection

Many Voices.
One Community.

INTELLIGENT PROTECTION

Pre-Filtering

Authentication protocols, reputation checks, threat intelligence feeds, and proprietary signatures.

AI-Powered

Natural language processing and AI/ML capabilities to analyze email content, identifying subtle anomalies and suspicious activity.

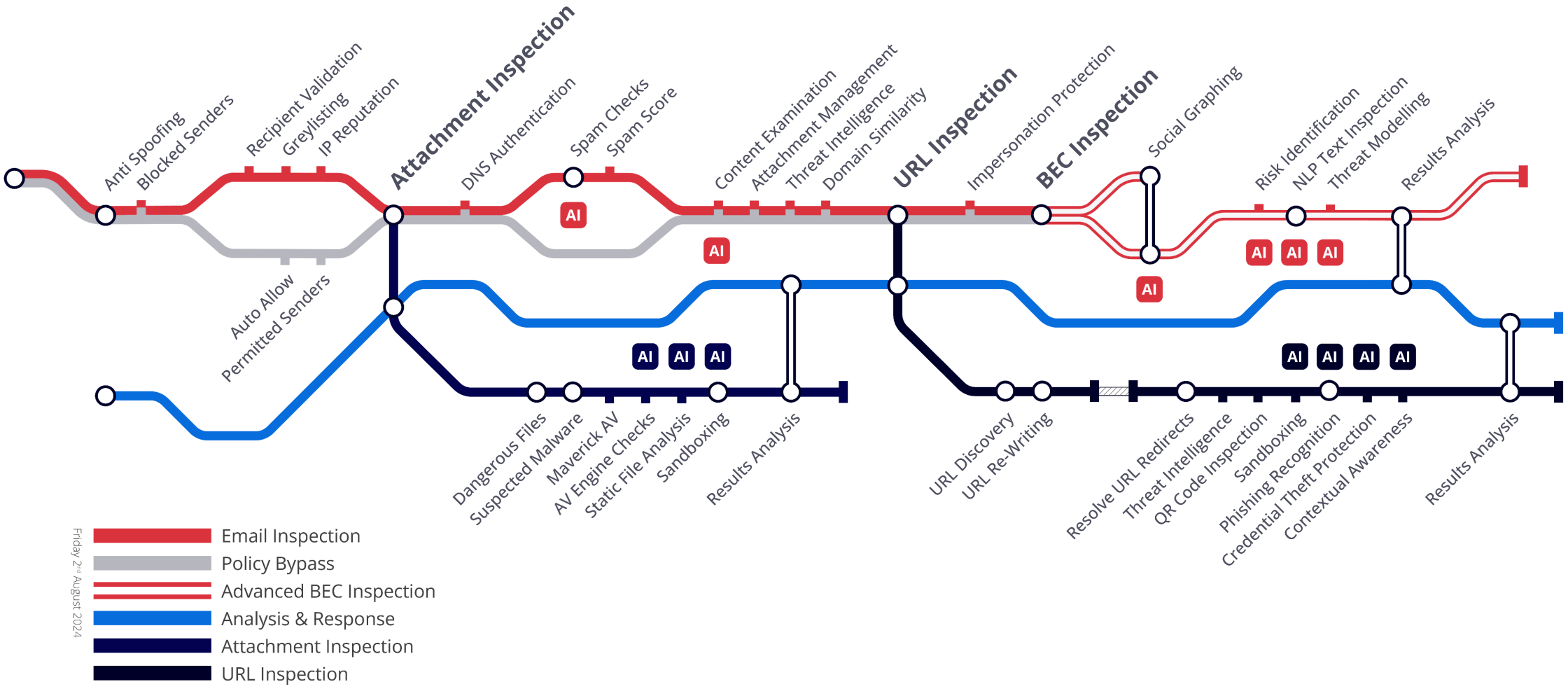
Integrated Detection

Multiple signals and technologies create a comprehensive view of threats, enabling blocking at the point of detection.

Continuous Learning

Billions of signals across the platform are analyzed to adapt and evolve ensuring highly effective detection against shifting tactics.

INTELLIGENT PROTECTION



BEC DETECTION

Header

Subject	Re: What time are you arriving at the office?
From Display	legalaw921@gmail.com Personal Email No Company Replies
From Envelope	Not available
Date/Time	01 Nov 2024 - 12:35:30
Message ID	<CACNcY6CKjBDH7Td+F7MXtb+wDwB463KQOeT+ziMDgP+MigNtUA@mail.gmail.com>

Body

I apologize for this inconvenience, I have a meeting and I will need you to
run an errand for Help with a Task me at a retail store close to you, I need some Apple Gift cards Gift Card . Do let me know if you can. Thanks.

On Tue, Sep 24, 2024 at 5:09 AM Courtney E. Jones <cjones@lawfirm.com> wrote:

> Should be in by 8
>
>
>
>
> COURTNEY JONES
> she|her|hers
> Attorney
> 555.682.459
> cjones@lawfirm.com
>

Header

Subject	Re: Task
From Display	No Company Replies
From Envelope	Not available
Date/Time	04 Nov 2024 - 12:53:22
Message ID	

Body

Zoe,

What are you up to at the moment? I'm scheduled for a meeting shortly,
while I also need to attend to a client. I need you to assist in getting a Help with a Task
task done at your end. Kindly spare a few minutes, and share with me your
available WhatsApp number Communication Channel Switch . This way, I could possibly have a chat while am in the meeting.

I look forward to your timely response.

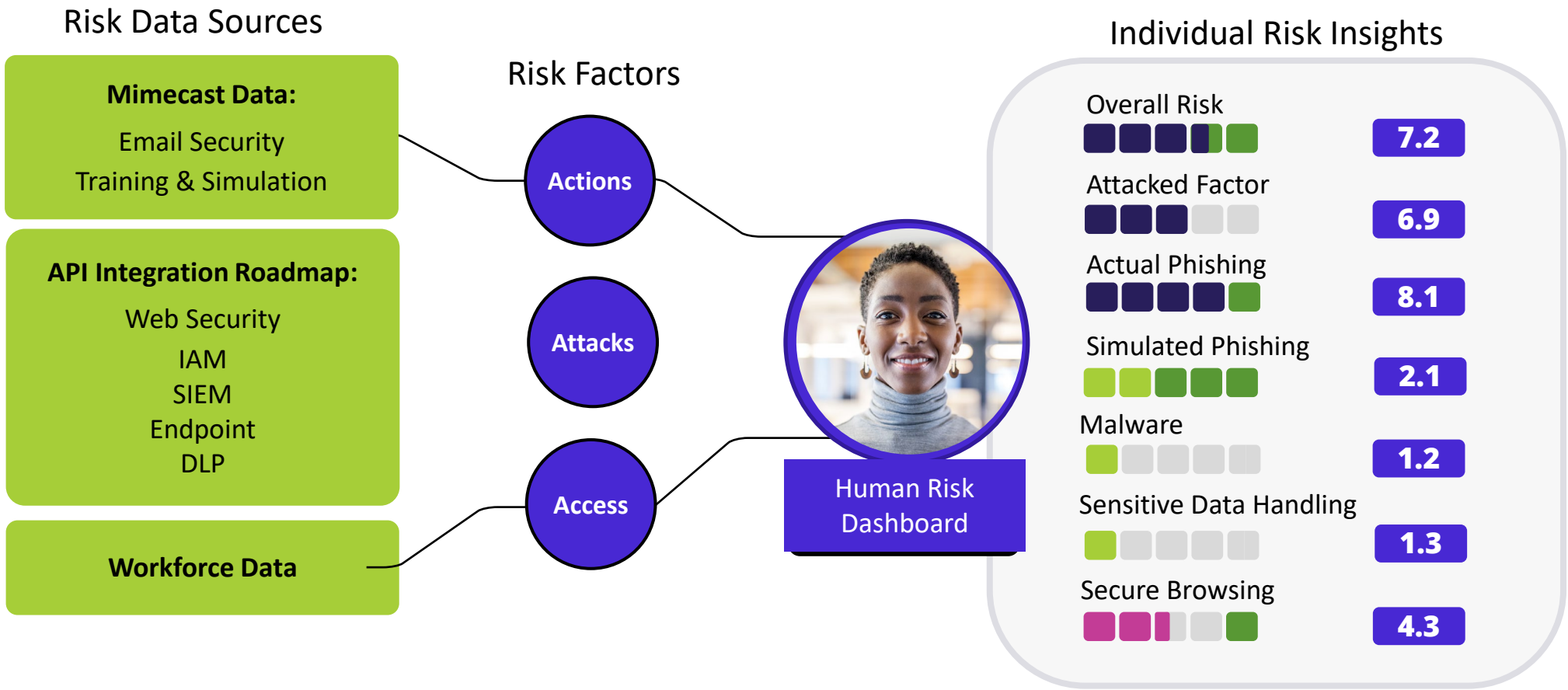
Kind regards,
Guy Bollen

Sent from my mobile

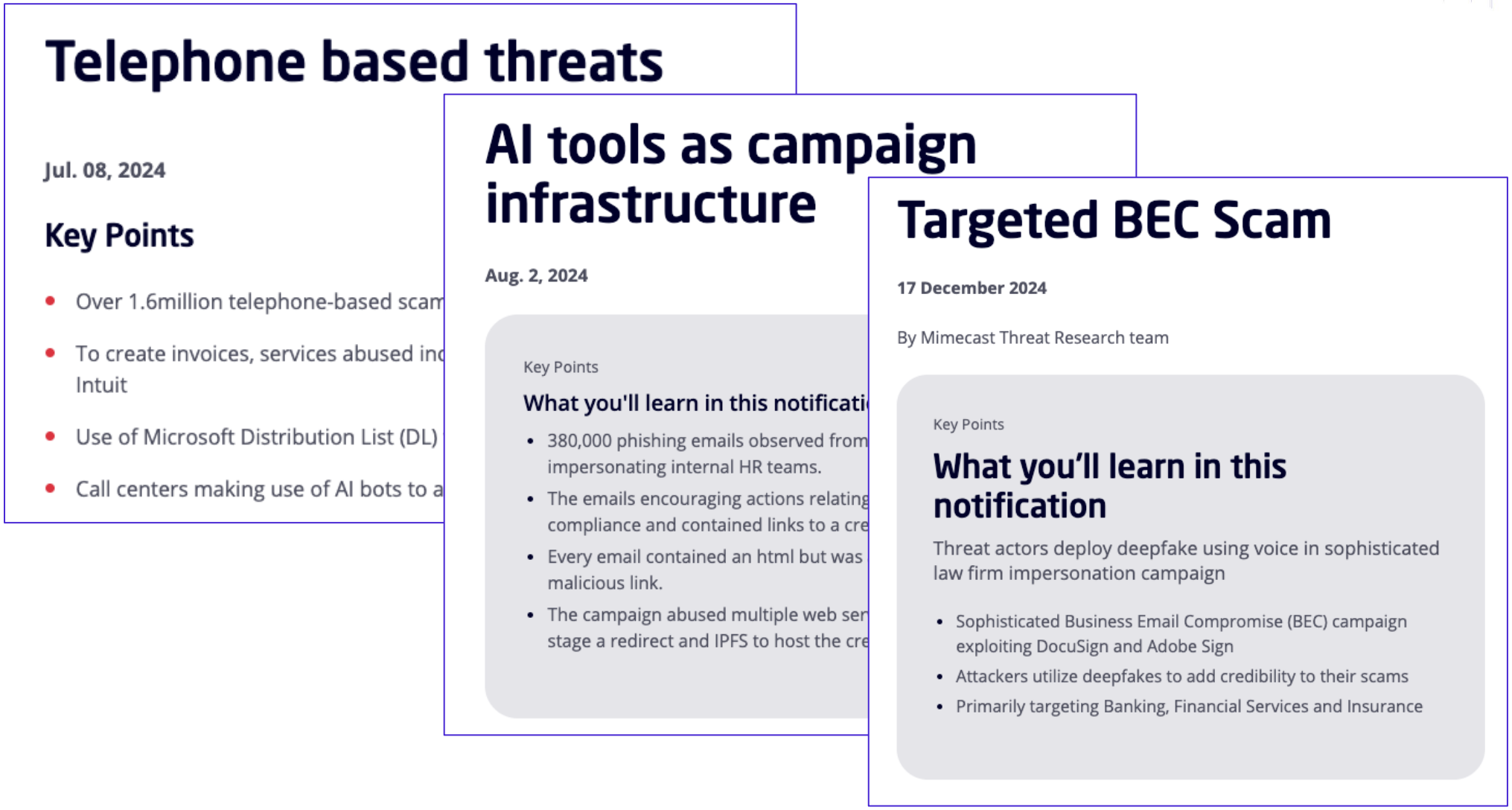
SECURITY AWARENESS, RE-ENVISIONED

Traditional Solutions		Powered by a Human Risk Management Platform
One-size-fits-all approach.		**Hyper personalized.**
Output oriented.		**Aligned with real security outcomes.**
Only measures simulated risk.		**Powered by real risk insights.**

HUMAN RISK DASHBOARD: UNPRECEDENTED VISIBILITY



BUT AI IS NOT JUST FOR EMAIL


Telephone based threats
Jul. 08, 2024
Key Points
Over 1.6million telephone-based scam
To create invoices, services abused inc
Intuit
Use of Microsoft Distribution List (DL)
Call centers making use of AI bots to a
AI tools as campaign infrastructure
Aug. 2, 2024
Key Points
What you'll learn in this notificati
380,000 phishing emails observed from
impersonating internal HR teams.
The emails encouraging actions relating
compliance and contained links to a cre
Every email contained an html but was
malicious link.
The campaign abused multiple web ser
stage a redirect and IPFS to host the cre
Targeted BEC Scam
17 December 2024
By Mimecast Threat Research team
Key Points
What you'll learn in this notification
Threat actors deploy deepfake using voice in sophisticated law firm impersonation campaign
Sophisticated Business Email Compromise (BEC) campaign exploiting DocuSign and Adobe Sign
Attackers utilize deepfakes to add credibility to their scams
Primarily targeting Banking, Financial Services and Insurance

RSAC™ 2025 Conference

Many Voices.
One Community.